INTRODUCTION TO THE DATABASE

The Chapman & Hall/CRC Chemical Database is a structured database holding information on chemical substances. It includes descriptive and numerical data on chemical, physical and biological properties of compounds; systematic and common names of compounds; literature references; structure diagrams and their associated connection tables. ***The Combined Chemical Dictionary on CD-ROM*** contains all those compounds published in:

- Dictionary of Organic Compounds (200,000 records)
- Dictionary of Natural Products (120,000 records)
- Dictionary of Inorganic and Organometallic Compounds (90,000 records)
- Dictionary of Pharmacological Agents (30,000 records)
- Dictionary of Analytical Reagents (14,000 records)

COMPOUND SELECTION

In general, CCD contains the following compounds:

- The basic fundamental organic and inorganic compounds of simple structure, including the elements, inorganic binary and ternary compounds (hydrides, halides, oxides, sulfides).
- Comprehensive coverage of virtually every known natural product including those of unknown structure.
- All currently marketed drugs, including all those listed in generic name compilations (US Adopted Names, International Nonproprietary Names, British Approved Names, Japanese Accepted Names), as well as those undergoing clinical trials.
- Compounds with an established use such as catalysts, solvents, starting materials, synthetic reagents, analytical reagents.
- Important coordination compounds, e.g. amines, phosphines, alkoxy complexes, and major well-characterised bioinorganics.
- Organometallic compounds representative of all important structural types (in the case of ligands with organic substituents, typically the parent member of each series, where known, together with a selection of homologues).
- Important biochemicals and minerals.
- Other compounds of particular interest because of their chemical, structural or biological properties, including many newly synthesised compounds of active research interest.

DATA PRESENTATION AND ORGANIZATION

Derivatives and variants

In the database, closely related compounds are grouped together to form an *entry*. Stereoisomers and derivatives of a parent compound are all listed under one entry. The compounds in ***The Combined Chemical Dictionary*** are grouped together into approximately 160,000 entries.

The structure of an entry is shown below.

Entry (parent compound)
Derivatives
Variants (stereoisomers or other closely-related compounds)
Derivatives of the variant

A simple entry covers one compound, with no derivatives or variants. A composite entry will start with the entry compound, then may have:

one or more derivatives at entry level

one or more variants of the entry

one or more derivatives of the variant.

Variants may include stereoisomers, e.g. (*R*)-form, *endo*-form; members of a series of natural products with closely related structures such as antibiotic complexes.

Molecular formulae are included for nearly all of these derivatives and so are readily searchable, whether they are documented as derivatives or have their own individual entry. Molecular formulae are not in general given for salts, hydrates or complexes (e.g. picrates) nor for most “characterisation” derivatives of carbonyl compounds such as 2,4-dinitro-phenylhydrazones and semicarbazones.

Where a derivative appears to have been characterised only as a salt, the properties of the salt may be given under the heading for the derivative. In such cases the data is clearly labelled, e.g. Mp 179° (as hydrochloride).

Where a compound can exist in several stereoisomeric forms, these isomers are grouped together into a single entry. Where known, physical properties, CAS registry numbers and pharmacological activity are assigned for each isomer.

Antibiotic complexes and biologicals

Many structurally related antibiotics co-occur as metabolites of the same producing organism, and these are presented within a single entry e.g. Actaplanin. Usually these components will only differ from each other in side chains or sugar residues and can be presented with a single structure diagram, suitably annotated.

Complex biological molecules such as hormones have been treated in a similar manner, with closely related species variants subsumed into a single entry. An example is Calcitonin.

Anions and cations

For ionic substances such as quaternary ammonium salts, the entry refers to the anion or cation, and the molecular formula and molecular weight given are those of the ion. The various salts (e.g. chloride, nitrate) are treated as derivatives, each with its own molecular formula.

Isotopically labelled compounds

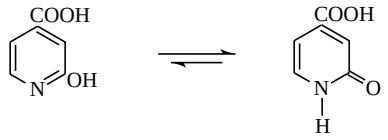
In general the coverage of CCD does not extend to isotopic variants. In cases where isotopes are used either medicinally or diagnostically, an entry for the unlabelled compound has been compiled and contains an explanation of the uses, plus additional CAS numbers applying to the isotopically labelled compound.

Minerals

Wherever possible, minerals corresponding to synthetic compounds are included within the entry for the synthetic compound, e.g. Zinc Blende and Wurtzite are incorporated into the entry for Zinc Sulfide (ZnS). Only when the mineral is the sole point of interest or unavailable synthetically is it given an entry in its own right.

Tautomerism

A completely consistent scheme for covering all such entries is not possible or desirable. Some variation is necessary in the way the appropriate CCD entries are organised in order to cover the various possibilities, but it is hoped and believed that the maximum possible clarity has been achieved. The general principles which have been followed are described here using as an illustration one of the commonest types of tautomerism exhibited by simple organic compounds, which is heterocyclic NH $\rightleftharpoons$ OH prototropy as exemplified by 2- and 4-hydroxypyridines.



2-Hydroxy-
4-pyridinecarboxylic acid
OH-form

1,2-Dihydro-
2-oxo-4-pyridinecarboxylic acid
NH-form

(*a*) Although in most simple cases the *NH*-form is the predominant tautomer in solution, the equilibrium is influenced by electronic and steric factors as the structure of the heterocycle varies. Entries often give a statement about tautomerism for a particular compound with supporting reference(s), but in many cases the individual compound will not have been studied closely and the probable tautomerism will have to be inferred.

(*b*) In the great majority of cases, the two (or more) tautomers and their derivatives are included in the same entry.

(*c*) All synonyms applicable to the tautomeric forms are given at the head of the entry.

(*d*) The entry name may refer to an unfavoured tautomer, for ease of presentation of a series of entries, e.g. in the above case the entry name is 2-Hydroxy-4-pyridinecarboxylic acid. Note that in such a series of isomers some (i.e. those with a 3-OH substituent) will not be capable of NH $\rightleftharpoons$ OH tautomerism of the type shown (although they may tautomerise to zwitterionic tautomers).

(*e*) For important compounds such as that shown above, structures are shown (or implied) for both tautomers. For less important compounds, the probable predominant tautomer may only be illustrated but synonyms are still given for all possible reasonable tautomers.

(*f*) Derivative data is given where appropriate under subheadings for the various tautomers. Derivatives may themselves be capable of tautomerism e.g. Me ester, (Methyl 2-hydroxy-4-pyridinecarboxylate), or may be blocked by substitution so that they clearly belong to one or other tautomer, e.g. Me ester (2-Methoxy-4-pyridinecarboxylic acid, derivative of the *OH*-form) and N–Me (derivative of the *NH*-form). They appear in the appropriate place in the entry.

(*g*) More complex examples where there are several possible tautomers not greatly differing in energy (e.g. purines, pteridines) are treated pragmatically to give the clearest possible presentation within the entry structure. The situation is complicated by the fact that some derivatives may be partially blocked and capable of tautomerism to fewer tautomeric structures than the parent. Such situations are usually covered by notes within the entry.

(*h*) CAS frequently indexes compounds where the tautomerism is unclear under a default structure, frequently the unfavoured 1*H*-form. If this is the case, a note is given in the entry.

Other very common types of tautomerism encountered include the

P(O)SH ⇌ P(S)OH interconversion shown by many organophosphorus compounds and the degenerate NH ⇌ NH tautomerism of pyrazoles and imidazoles. The same general principles have been followed and should be clear from inspection of the individual entries. For some types of organophosphorus tautomerism, e.g. phosphinic acids RPH(O)OH ⇌ phosphonous acids $RP(OH)_2$, separate entries for the two substances have been retained.

DATA TYPES

The format of a typical entry is given in Fig. 1, and shows the individual types of data that may be present in an entry.

Chemical names and synonyms

All of the names discussed below can be searched using the Chemical Name field.

The Entry Name is that chosen to head each entry and is that which, in the opinion of the Editors, is most likely to be known by, and of use to, most readers. Systematic Names following IUPAC conventions are used wherever convenient, but trivial names may be used for more complex structures such as pharmaceuticals and natural products. In cases where no one name stands out as being clearly more familiar or convenient than others, the Chemical Abstracts name is normally used as the entry name.

For a fuller treatment of nomenclature principles and details, see *The Organic Chemist's Desk Reference* (Chapman & Hall, 1995).

An important function of CCD is to present a wide range of synonyms. In general, the selection is made as useful as possible, but no attempt is made to provide exhaustive lists of proprietary names for pharmaceuticals and other commercial substances.

Archaic systematic names are in general not given, but obsolete synonyms have often been retained where there has been a change in numbering of the parent ring system and these synonyms could assist readers who have to consult the older literature. In a few cases incorrect synonyms from the literature have also been reported. Synonyms in these classes are distinguished as '*obsol*' or '*incorrect*' respectively. Several obsolescent systems such as the carbinol and hydroxyalkane alcohol nomenclatures have been almost completely discarded, since although they are still occasionally met with, users should have no difficulty in converting these to the normal nomenclature.

Names which are known to be duplicated within the chemical literature are marked with the sign †. These are usually duplicate trivial names for natural products or pharmaceuticals, but there are a few cases (of organophosphorus compounds) where two or more compounds of different structure have been allocated the same CAS name.

CAS names. Names corresponding to those used by CAS during the 8th through to the 12th Collective Index Periods (1967–71, 1972–76, 1977–81, 1982–86, 1987–1991 respectively) are labelled with the suffixes 8CI, 9CI, 10CI, 11CI and 12CI respectively. Names encountered in CAS since 1991 are labelled 13CI although it is possible that some further changes may have occurred before publication of the 13th Collective Index.

For the majority of organic and inorganic compounds, and simple organometallic compounds such as metal alkyls, the nomenclature brought in for the 9th Collective Index Period (and referred to as 9CI nomenclature) has

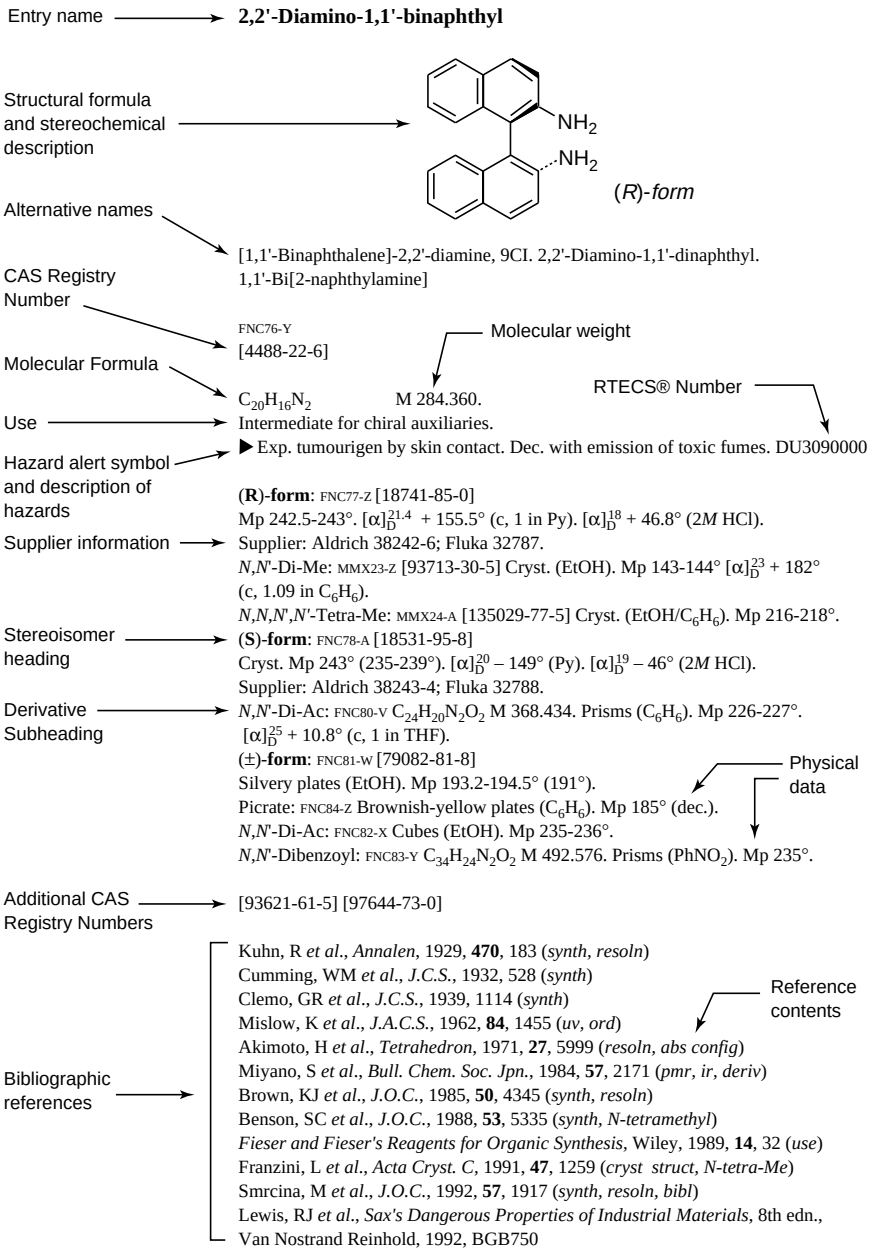
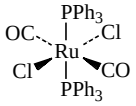


Fig. 1. Sample entry from database

since been unchanged. This is not true for some groups of compounds such as cluster boranes and the more complex organometallic compounds, where the nomenclature is still evolving. There are also many examples of the same compound being registered more than once under different names (and registry numbers) in CAS.

The following types of suffix which are to be found attached to CAS names have been omitted. Firstly, stereochemical descriptors, e.g. in Dicarbonyldichloro-bis(triphenylphosphine)ruthenium the CAS descriptor (*OC*-6-12) indicates the geometry shown below:



In CCD this is referred to as the *af*-dicarbonyl-*bd*-dichloro-*ce*-diphosphine form. Secondly, bonding descriptors, e.g. in Hexa-μ-chlorohexachlorotriruthenate(4-) the CAS descriptor (2*Ru*-*Ru*) denotes the presence of two ruthenium-ruthenium bonds. On the other hand, oxidation state has in many cases been inserted in CAS names.

Oxidation states and charges. For any given substance, the oxidation state (also known as the Stock number) of the element of interest is incorporated into at least one of the names given, using Roman numerals or zero, provided that it can be unequivocally assigned. Oxidation states are therefore generally omitted from nitrosyl complexes where the assignment of oxidation state is often controversial, and also from compounds of elements having only one common oxidation state, where it is unnecessary.

The overall ionic charge (also known as the Ewens-Bassett number) of a complex is also provided in at least one name, using Arabic numerals.

CAS names do not describe oxidation states, only charges. However, where the CAS name is the only readily accessible one, the oxidation state has been added editorially. Where both oxidation state and charge occur in a single name, the former precedes the latter.

Spelling conventions

American spelling is generally used for chemical names:

- sulf- not sulph
- estr- not oestr-

with some exceptions:

- caesium not cesium
- aluminium not aluminum

UK spelling is used throughout the text.

CAS Registry Numbers

CAS Registry Numbers are identifying numbers allocated to each distinctly definable chemical substance indexed by the Chemical Abstracts Service since 1965 (plus retrospective allocation of numbers by CAS to compounds from the sixth and seventh Collective Index Periods). The numbers have no chemical significance but they provide a label for each substance independent of any system of nomenclature.

Much effort has been expended to ensure that accurate CAS numbers are given for as many substances as possible. If a CAS number is not given for a particular compound, it may be (*a*) because CAS have not allocated one, (*b*) very occasionally, because an editorial decision cannot be made as to the correct number to cite, or (*c*) because the substance was added to the

database at a late stage in the compilation process, in which case the number will probably be added to the database soon.

At the foot of the entry, immediately before the references, may be shown additional registry numbers. These are numbers which have been recognised by the Editors or contributors as belonging to the entry concerned but which cannot be unequivocally assigned to any of the compounds covered by the entry. Their main use will be in helping those who need to carry out additional searches, especially online searches in the CAS or other databases, and who will be able to obtain additional hits using these numbers. Clearly, discretion is needed in their use for this purpose.

Additional registry numbers may arise for a variety of reasons:

(*a*) A CAS number may refer to stereoisomers or other variants of the main entry compound, e.g. bonding isomers, for which no physical properties or useful information is available. In many cases, although CAS numbers are allocated to different isomers, they are not assigned specifically to each one and are merely labelled ‘stereoisomers’.

(*b*) Hydrates, salts, complexes, etc. which are not characterised fully.

(*c*) A CAS number may refer to a mixture or to a particular non-stoichiometric composition which is not detailed individually in the entry.

(*d*) Replaced numbers, duplicate numbers and other numbers arising from CAS indexing procedure or, occasionally, from errors or inconsistences by CAS, are also reported.

Diagrams

In each entry display there is a single diagram which applies to the parent entry. Separate diagrams are not given for variants or derivatives.

Every attempt has been made to present the structures of chemical substances as accurately as possible according to current best practice and IUPAC recommendations. In drawing the formulae, as much consistency as possible between closely related structures has been aimed at. Thus, for example, sugars have been standardised as Haworth formulae and, wherever possible in complex structures, the rings are oriented in the standard Haworth manner so that structural comparisons can quickly be made. In formulae the pseudoatom abbreviations Me, Et and Ac for methyl, ethyl and acetyl respectively, are generally used when attached to a heteroatom. Ph is used throughout whether attached to carbon or to a heteroatom. Other pseudoatom abbreviations such as Pr^i for isopropyl and Bz for benzoyl are not used.

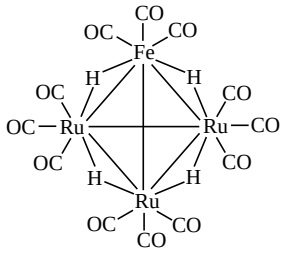
Peptides are usually drawn using the standard three letter abbreviations. The configurations of constituent amino acids are assumed to be L- unless otherwise stated.

Care must be taken with the numbering of natural products, as problems may arise due to differences in systematic and non-systematic schemes. Biogenetic numbering schemes which are generally favoured in CCD may not always be contiguous, e.g., where one or more carbon atoms have been lost during biogenesis.

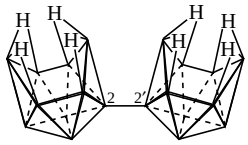
Structures for derivatives can be viewed in **Structure Search**, but remember that these structures are generated from connection tables and may not always be oriented consistently.

Bonding in inorganic and organometallic compounds. The bonding in many transition metal complexes and clusters is more or less complex and subject to varying interpretation, and is therefore not amenable to accurate depiction by the conventions which serve reasonably well for organic compounds.

Bridging hydrogens between two metal centres are depicted for clarity as though there are full metal-metal bonds, although there is rarely so much electron density between the two metals, e.g.:



Boranes. BH groups in the cluster boranes and related species are represented by vertices, as shown below:



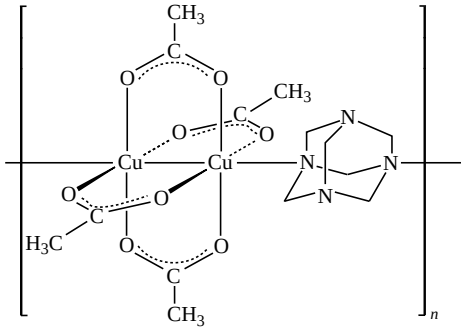
Only when B is bonded to 2 (or more) non-bridging atoms is it depicted explicitly. All other atoms, including carbon, are depicted explicitly.

This convention is analogous to the representation of CH_2 or CH groups as plain vertices in organic compounds and which is also used to depict ligands in this database.

Transition metal complexes. Very considerable variations in conventions for depicting organometallic compounds are to be found in the literature. The bonding situation in many transition metal complexes is more or less complex and subject to varying interpretations, and is therefore not amenable to accurate depiction by the conventions which serve reasonably well for organic compounds. For example, the two following representations of the complex obtained from octacarbonyldicobalt and acetylene refer to the same compound:

For sandwich complexes, the following convention, illustrated with ferrocene as an example, is used throughout:

Polymeric transition metal complexes. Wherever possible, the coordination polyhedron of the metal is depicted, and the points of attachment to the next unit are indicated using bonds that extend outside square brackets, e.g.



Stereochemical conventions

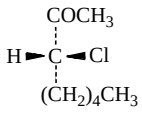
Where the absolute configuration of a compound is known or can be inferred from the published literature without undue difficulty, this is indicated. Where only one stereoisomer is referred to in the text, the structural diagram indicates that stereoisomer. Wherever possible, stereostructures are described using the Cahn-Ingold-Prelog sequence-rule (*R,S)* and *(E,Z)* conventions but, in cases where these are cumbersome or inapplicable, alternatives such as the α,β-system are used instead. Alternative designations are frequently presented in such cases.

(a) *The* (*R,S*)*-system*

In the simplest case, the four substituent atoms about a tetrahedral carbon atom are placed in order of decreasing atomic number and the molecule is then viewed from the side remote from the substituent of lowest priority. The configuration is *(R) (rectus)* if the order of the three other groups from highest to lowest is clockwise, and *(S) (sinister)* if it is anticlockwise.

(*R*) (*S*)

If two or more of the four atoms attached to the central atom are identical, the molecule is explored outwards by a process of comparing atom with atom.

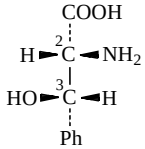


(*R*)-3-Chloro-2-octanone
(priority Cl> C═O>····CH_2····

Extensions of the *(R,S)*-system refer to situations such as axial and planar chirality (biaryls, cyclophanes, etc.) and to molecules with central atoms other than carbon (e.g. chiral sulfoxides).

Where only the relative configuration of a compound containing more than one chiral centre is known, the symbols (*R**) and (*S**) are used, the lowest-numbered chiral centre being arbitrarily assigned the symbol (*R**). For racemic modifications of compounds containing more than one chiral centre the symbols (*RS*) and (*SR*) are used, with the lowest-numbered chiral centre being arbitrarily assigned the symbol (*RS*). The racemate of a compound containing one chiral centre only is described in CCD as (±)-.

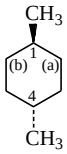
In comparing CAS descriptors with those given in CCD, it is important to remember that the order of presentation of the chirality labels in CAS is itself based on the sequence rule priority and not on any numbering scheme. For example in CCD, the following compound:



is (2*R*,3*S*)-2-Amino-3-hydroxy-phenylpropanoic acid. In CAS it is [*S*-(*R**,*S**)]-β- Hydroxyphenylalanine. The relative stereochemical label (*R**,*S**) is first applied with the *R* applying to chiral centre 3* because it has higher priority than centre 2 (OH > NH_2). The absolute stereochemical descriptor (*S*)- is then applied changing *R** to *S* for chiral centre 3 and *S** to *R* for chiral centre 2. For further details, see the current CAS Index Guide.

For simplicity, the enantiomers of bridged-ring compounds, such as camphor, are described simply as (+)- and (–)-. Although camphor has two chiral centres, steric restraints mean that only one pair of enantiomers can be prepared.

The (*R*,*S*) descriptor system can be extended to describe the configurations of many types of symmetrical compound, e.g. the 1,4-Dimethylcyclohexanes.



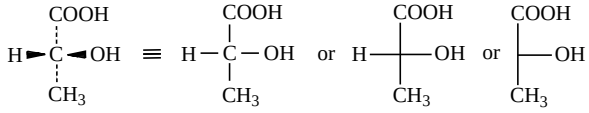
(1*RS*,4*RS*) 1,4-Dimethylcyclohexane

At chiral centre 1, an arbitrary choice is made between the two equivalent sequence chains (a) and (b). Choosing (a) arbitrarily gives (a) > (b) > CH_3 > H at C(1) leading to (*R*)-configuration. The chirality at C(4) is then (a) > (b) > CH_3 > H or (*R*-). The configuration of the compound can therefore be described as (1*RS*,4*RS*) (or 1*R**,4*R**). This is independent of the arbitrary choice made.

For further information on the (*R*,*S*)-system, see Cahn, R,S *et al, J. Chem. Soc.*, 1951, 612; *Experientia,* 1956, **12**, 81; *Angew. Chem. Int. Ed. Engl.*, 1966, **5**, 383.

Where appropriate, alternative stereochemical descriptors may be given using the D, L or α,β-systems. For a fuller description of these systems, consult *The Organic Chemist's Desk Reference*.

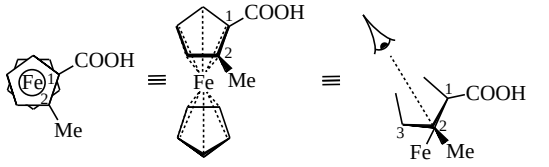
The structure diagrams for compounds containing one or two chiral centres are given in CCD as Fischer-*type* diagrams showing the stereochemistry unequivocally. True Fischer diagrams in which the configuration is implied by the North-South-East-West positions of the substituents are widespread in the literature; they are quite unambiguous but need to be used with caution by the inexperienced. They cannot be reoriented without the risk of introducing errors.



CCD representation True Fischer projections

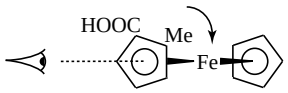
The use of the (*R*,*S*)-system for chiral polyhapto complexes is not covered by the original Cahn-Ingold-Prelog rules and further specification of ligand priorities and bonding convention is required.

Chiral metallocenes and related complexes. The most widely employed system for specification of metallocene chirality is due to Schlögl. The bond from the central metal atom to the ring carbon atom under consideration is treated as a formal single bond. The carbon atom is then considered as a chiral centre and (*R*,*S*) nomenclature is applied in the usual way.



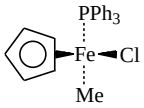
(1*S*,2*R*) 2-Methylferrocene 1-carboxylic acid Priority Fe > C_1 > C_3 > Me chirality (2*R*)

For further information see Schlögl, K., *Topics in Stereochemistry*, 1967, 39. In some older papers, the molecule is considered overall as a case of planar chirality. However, this convention becomes ambiguous when applied to some more complex structures.



$(R)_{planer}$

Polyhapto ligand as a substituent on a chiral atom. Several conventions have been proposed for determining the order of priority of ligands where one or more is π-bonded. Probably the one most widely accepted is due to Stanley and Baird, in which the ligand is considered a pseudoatom of atomic weight equal to the sum of all of the π-bonded atoms.

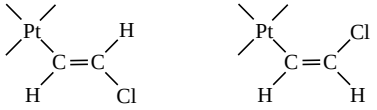


Priority C_5H_5 ('atomic weight' = 60) > Cl > PPh_3 > Me
Chirality (*S*) at Fe

For further information see Stanley, K. *et al*, *J. Am. Chem. Soc*., 1975, **97**, 6598.

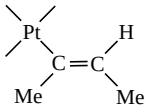
(b) *The (E,Z)-system*

This is an extension of the (*R,S*)-system for specifying configurations at alkene double bonds. The substituents are ordered as in the (*R,S*)-system and if the two of higher priority are on the same side of the double bond, the configuration is (*Z*) (*zusammen*), while if they are on opposite sides it is (*E*) (*entgegen*).



(*E*) (*Z*)

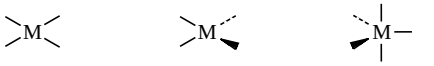
Note that (*E*) does not always correspond to the *trans*- of the earlier literature.



cis (with respect to Me groups)
(*E*) (priority Pt > Me; Me > H)

(c) *Coordination polyhedra*

The various coordination polyhedra are depicted using wedged and dashed bonds, the most common polyhedra being:



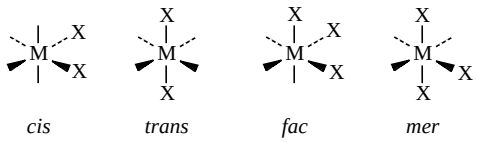
Square planar Tetrahedral Trigonal bipyramidal

Square pyramidal Octahedral

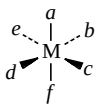
The shapes of polyhedra greater than 6 are not amenable to clear representation by this means and a textual statement such as ‘square antiprismatic’ is combined with the diagram.

The terms ‘tetrahedral’ and ‘octahedral’ are used in a general sense and do not imply strict symmetry types. For the latter, the point group descriptors T_d and O_h are employed.

In the case of **octahedral complexes** bearing two different types of substituents, the stereochemistry is adequately defined using the terms *cis*, *trans*, *fac* or *mer*:



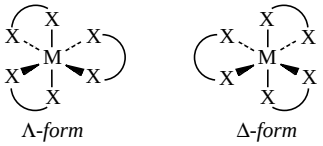
In more complicated cases, italicised letters are used to designate the positions of ligands in various configurations. The letters are assigned thus:



The first mentioned alphabetical ligand in the name is given the designator *a*, the second ligand the next lowest designator and the assignments to the remaining ligands then follow from this.

Stereochemistry for polydentate ligands is described using the α and β convention:

The absolute configuration of certain octahedral complexes is described using the Δ, Λ convention:



Molecular formula and molecular weight

The elements in the molecular formula are given according to the Hill convention (C, H, then other elements in alphabetical order). The molecular weights given are formula weights (or more strictly, molar masses in daltons) and are rounded to one place of decimals. In the case of some high molecular mass substances such as proteins the value quoted may be that taken from an original literature source and may be an aggregate molar mass.

Source/Synthesis

The biological source is quoted for all naturally occurring compounds. The taxonomic names for organisms given throughout are in general those given in the primary literature. Standardisation of minor orthographical variations has been carried out. Data in this field may be searched under **Source/Synthesis** or **All Text**. Standards used are: Brummitt, R.K. (1992) *Vascular Plant Families and Genera*, Royal Botanic Gardens, Kew; Willis,

J.C. (1973) *A Dictionary of the Flowering Plants*, Cambridge University Press, Cambridge; Gozmany, L. (1990) *Seven Language Thesaurus of European Animals*, Chapman & Hall London; Chemical Abstracts Service.

Brief synthetic details are provided for the majority of inorganic and organometallic compounds, with the aim of apprising the user with the kind of routes available. Formulae are frequently abbreviated to save space (the most common ones are included in Table 1. For full details of conditions necessary to effect transformations, the user is referred to the appropriate literature citations provided. Information about the sources of compounds occurring as minerals or natural products is also given.

Importance/use

Care has been taken to make the information given on the importance and uses of chemical substances as accurate as possible. Many substances have now been patented for a wide variety of uses but this does not imply that the patented uses are of widespread applicability or even of established utility. In general, information on a particular use is given prominence only when it is documented in a critical source, such as *Kirk-Othmer* or *Ullmann*, when it is protected by numerous patents, or when a reference is quoted which will assist the reader to assess the value of the claimed application. Data in this field may be searched under **Use/Importance** or **All Text**.

Use of organometallic compounds as synthetic reagents is now widespread and this is reflected in the addition of Synthetic Reagents Classification Codes, which are searchable under the **Type of Compound** field.

Practically all pharmacologically active compounds have been classified according to its mechanism of action (where this is known) and therapeutic use, under one or more headings in the **Type of Compound** field. Compounds which have several claimed therapeutic uses are indexed separately in each category.

Physical Data

Interatomic Distances

Selected dimensions for inorganic and organometallic compounds usually obtained by x-ray crystallography, are provided: bond lengths are given in picometres (pm = 10^{-12}m = 10^{-2}Å) and angles in degrees.

Appearance

This data describes whether a compound is solid, liquid or gas and also gives an indication of its colour (even if colourless), crystal form and recrystallisation solvent. Details of air, moisture and thermal stability are also included where available.

Melting points and boiling points

These are reported in degrees Celsius. The policy followed in cases of conflicting data is as follows:

- Where the literature melting points are very similar, only one figure (the highest or most probable) is quoted
- Where two or more melting points are recorded and differ by several degrees (the most likely explanation being that one sample was impure), the lower figure is given in parentheses; thus Mp 139° (135–136°).
- Where quoted figures differ widely and some other explanation such as polymorphism or incorrect identity seems the most likely explanation, both figures are quoted without parentheses; thus Mp 142°, Mp 205–206°.

- Known cases of polymorphism or double melting point are noted.
- Boiling points are given at atmospheric pressure unless otherwise indicated. The pressure in mmHg (if not atmospheric) is given as a subscript, e.g. Bp_{100} 85°. Some boiling points are now quoted in the literature with the pressure in kilopascals (kPa, SI units). The conversion factor is; 1 mmHg = 0.133222 kPa; 1 kPa = 7.50064 mmHg.
- Boiling point determination is less precise than that of melting points and conflicting boiling point data is not usually reported except when there appears to be a serious discrepancy between different authors.
- Sublimation temperatures are recorded in a similar style to boiling points e.g. $Subl._{20}$ 130°.

Optical rotations

These are given wherever possible and are expressed in the form: $[\alpha]_D^{20}$ + 30.6 (c, 1.2 in MeOH). This denotes a temperature of 20°C, wavelength at the sodium D line (589 nm) and a concentration of 1.2 g/100 ml in methanol solution. Where reported in the literature, an indication of optical purity (op) or enantiomeric excess (ee) is recorded after the rotation value.

The degree sign following optical rotations, although still extensively found in the primary literature, has been dropped as it is dimensionally incorrect.

Densities and refractive indices

Although these are now of less importance for the identification of liquids than has been the case in the past, they are still quoted for common substances.

Densities and refractive indices are not quoted where the determination appears to refer to an undefined mixture of stereoisomers.

Spectroscopic data

Spectroscopic data such as ir maxima, uv wavelengths and extinction coefficients are given in many cases where spectroscopic identification has been important in characterisation, particularly for unstable compounds. Efforts have been made, in particular, to include carbonyl and M—H stretching frequencies wherever possible. In many other cases, spectroscopic data can be rapidly located through the references quoted.

Thermodynamic data

Limited thermodynamic data is provided. In many other cases, this information can be located through the references quoted.

pK_a values

Experimentally determined pK_a values are given for both acids and bases. The pK_a of a base can be obtained by subtracting its pK_b from 14.17 (at 20°C) or from 14.00 (at 25°C).

Octanol-water partition coefficients

Octanol-water partition coefficients are a useful measure of a compound's lipophilicity, and hence its potential biodistribution. Calculated values have been included for as many pharmacologically active compounds as possible, and are denoted by the suffix 'calc' in parentheses. These have been calculated by Tripos, Inc.† using the Hansch and Leo algorithm CLOGP.

† TRIPOS, Inc., 1699 S. Hanley Road, St Louis, MO, USA. Phone: 314 647 1099, 800 323 2960. Fax: 314 647 9241.

Development status

This field gives information (when available) about whether a compound has been marketed, or is undergoing clinical trials. If a compound has been withdrawn from the market, this is also stated. Although a variety of sources have been consulted, the transient nature of this type of information means that this dictionary should not be considered as an authoritative or up to date source for a drug's marketing status. Details of the worldwide top 100 prescription drugs (in 1995) are also provided.

Hazard and toxicity information

General

Toxicity and hazard information is displayed in red type and additionally is highlighted by the sign ▶ . It has been selected to assist in risk assessments for experimental, manufacturing and manipulative procedures with chemicals.

Physical, reactive and ***toxic*** properties all contribute to the hazard associated with a particular chemical. As part of the ***physical data***, flash points, explosive limits and autoignition temperatures have been included (where appropriate). Flammability classifications, which are based on flash point measurements and boiling points, are also mentioned, and the opportunity has been taken to include UK occupational exposure limits, or for some compounds threshold limit values published by the American Conference of Governmental Industrial Hygienists (ACGIH).

For the ***reactive*** hazards, a brief comment is made on any explosive (or violent polymerisation) properties and aspects of the chemical reactivity of a substance which are of concern. These include the potential for peroxidation, oxidizing/reducing properties and incompatibility with commonly available chemicals.

Toxicity information has been chosen to show hazardous effects from short-term or long-term exposure. Observations from human exposure are summarised if available (including possible adverse effects of drugs), otherwise experimental (exp.) tests are quoted. Included in the toxicity data are the results of irritancy tests, acute lethality data, target organ toxicity, and carcinogenic and reproductive properties where appropriate. Those chemicals which have been classified by the International Agency for Research on Cancer (IARC) as *human carcinogens*, *probable human carcinogens* or *possible human carcinogens* have been identified accordingly.

Many inorganic and organometallic compounds have not been evaluated toxicologically but it is to be assumed that all compounds of certain elements such as As, Be, Hg and Tl are toxic, and that compounds containing certain groups such as perchlorate and azide are likely to be explosive.

The handling of the majority of air-sensitive inorganic and organometallic compounds is to be regarded as hazardous to a greater of lesser degree because of the risk of fire or explosion in contact with air. Not every such sensitive compound has been specially marked as hazardous. Additionally, many metal halides (often the starting point for organometallic synthesis) can be easily hydrolysed, and all should be regarded as skin, eye and respiratory tract irritants.

The Publishers cannot be held responsible for any inaccuracies in the reported information, neither does the omission of hazard data in the ***Dictionary*** imply an absence of this data from the literature. Widely recognised hazards are included however, and where possible key toxicity reviews are identified in the references. Further advice on the storage,

handling and disposal of chemicals is given in *The Organic Chemist's Desk Reference*.

Finally, it should be emphasised that any chemical has the potential for harm if it is carelessly used. For many newly synthesised materials (e.g. new synthetic reagents), hazardous properties may not be apparent or may have been cited in the literature. In addition, the toxicity of some very reactive chemicals may not have been evaluated for ethical reasons, and these substances in particular should be handled with caution.

*RTECS® Accession Numbers**

Many entries in DOC contain one or more RTECS® Accession Numbers. Possession of these numbers allows users to locate toxicity information on relevant substances from the NIOSH *Registry of Toxic Effects of Chemical Substances*, which is a compendium of toxicity data extracted from the scientific literature.

For each Accession Number, the RTECS® database provides the following data when available: substance prime name and synonyms; date when the substance record was last updated; CAS Registry Number; molecular weight and formula; reproductive, tumorigenic, and toxic dose data; and citations to aquatic toxicity ratings, IARC reviews, ACGIH Threshold Limit Values, toxicological reviews, existing Federal standards, the NIOSH criteria document program for recommended standards, the NIOSH current intelligence program, the NCI Carcinogenesis Testing Program, and the EPA Toxic Substances Control Act inventory. Each data line and citation is referenced to the source from which the information was extracted.

Bibliographic References

The selection of references is made with the aim of facilitating entry into the literature for the user who wishes to locate more detailed information about a particular compound. Thus, in general, recent references are preferred to older ones, particularly for chiral compounds where optical purity and absolute configuration may have been determined relatively recently. The number of references quoted cannot therefore be taken as an indication of the relative importance of a compound, and the references quoted for important substances may not be the most significant historically. For very common compounds which are nowadays readily available from bulk suppliers, long lists of syntheses are not presented, but the emphasis is on references to spectra, chromatography, etc.

References are given in date order except for references to spectroscopic library collections, which sort at the top of the list, and those to hazard/toxicity sources which sort at the bottom.

The content of many references are indicated by means of suffixes, or tags. A list of the most common ones is given in Table 2.

Some reference suffixes are now given in **boldface** type, where the editors consider the reference to be particularly important, for example the best synthesis giving full experimental details and often claiming a higher yield than previously reported methods.

For many compounds, especially those still undergoing clinical trials, the only information available may be in the patent literature. Wherever possible the original patents (or English language equivalents) have been consulted, and details of biological activity and physical properties are as quoted in these documents. When referring to a patent no distinction is made between patent applications and granted patents.

As well as quoting the patent office and patent number, a CAS reference and the patenting company are also given. Patents are usually quoted in the following format:

Eur. Pat., 1987, Glaxo, 247 458; CA, **116**, 7887q (synth)

In some entries, minor items of information, particularly the physical properties of derivatives, may arise from references not cited in the entry.

Journal abbreviations

In general these are uniform with the *Chemical Abstracts Service Source Index* (CASSI) listing except for a short list of very common journals:

CCD ABBREVIATION	***CASSI***
Acta Cryst. (and sections thereof)	Acta Crystallogr. (and sections thereof)
Annalen	Justus Liebigs Ann. Chem.
Chem. Comm.	J. Chem. Soc., Chem. Commun.
J.A.C.S.	J. Am. Chem. Soc.
J.C.S. (and various subsections thereof)	J. Chem. Soc. (and various subsections thereof)
J. Het. Chem.	J. Heterocycl. Chem.
J.O.C.	J. Org. Chem.
Tet. Lett.	Tetrahedron Lett.

Table 1. Database abbreviations

Abbreviation	***Meaning***
[α]	specific rotation
acac	acetylacetonato
Ac	acetyl
ACGIH	American Conference of Governmental Industrial Hygienists
Ac_2O	acetic anhydride
AcOH	acetic acid
ADI	Acceptable Daily Intake
alk.	alkaline
amorph.	amorphous
ANSI	American National Standards Institute
anhyd.	anhydrous
approx.	approximately
aq.	aqueous
asym.	asymmetrical, unsymmetrical
B	base
BAN	British Approved Name
biol.	biological
bipy	2,2′-bipyridine
Bp	boiling point
br	broad
BSI	British Standards Institution
Bu	butyl (Bu^t for *tert*-butyl etc.)
bwd	bird (wild)
Bz	benzyl
c.	concentration
ca.	(circa) about

Abbreviation	*Meaning*
CAS	Chemical Abstracts Service
ccp	cubic close packed
cdt	1,5,9-cyclododecatriene
C_6H_6	benzene
C_5Me_5	pentamethylcyclopentadienyl
CNS	central nervous system
cod	1,5-cyclooctadiene
col.	colour, coloration
comly.	commercially
compd(s)	compounds(s)
conc.	concentrated
const.	constant
constit.	constituent
coord	coordinate(d), coordination
cot	1,3,5,7-cyclooctatetraene
Cp	cyclopentadienyl
C_5Ph_5	pentaphenylcyclopentadienyl
cryst.	crystal(s)
cv	cultivar
CVD	chemical vapour deposition
Cy	cyclohexyl
d	density
dba	dibenzylideneacetone
dck	duck
dec.	decomposes, decomposition
degradn.	degradation
depe	1,2-bis(diethylphosphino)ethane
descr.	described
diars	diarsine (generalised ligand)
dil.	dilute, dilution
dimorph.	dimorphic
diphos	diphosphine (generalised ligand)
diss.	dissolves, dissolved
dissoc.	dissociates
dist.	distil, distillation
DMA	dimethylacetamide
DMF	dimethylformamide
dmpe	1,2-bis(dimethylphosphino)ethane
dmpm	bis(dimethylphosphino)methane
DMSO	dimethyl sulfoxide
dppe	1,2-bis(diphenylphosphino)ethane
dppm	bis(diphenylphosphino)methane
dppp	1,3-bis(diphenylphosphino)propane
EDTA	ethylenediaminetetracetate(4-)
ee	enantiomeric excess
E_g	band gap (electron volts)
en	ethylenediamine
equilib.	equilibrium
esp.	especially
Et	ethyl
EtOAc	ethyl acetate
EtOH	ethanol
EtOH aq.	aqueous ethanol
evapn.	evaporation
exp.	exposure

Abbreviation	*Meaning*
exp.	experimental
fac	facial
Fc	ferrocenyl
fl. p.	flash point
fluor.	fluoresces, fluorescence
formn.	formation
Fp	freezing point
g	gram(s)
ΔG_f^0	standard free energy of formation
Glc	β-D-glucopyranosyl
gpg	guinea pig
ham	hamster
ΔH_f^0	standard enthalpy of formation
hcp	hexagonal close packed
hydrol.	hydrolyses, hydrolysed, hydrolysis
ihl	inhalation
im	imidazolato
ims	intramuscular
INN	International Non-proprietary Name
inorg.	inorganic
insol.	insoluble
intermed.	intermediate
ipr	intraperitoneal
ISO	International Standards Organisation
ivg	intravaginal
ivn	intravenous
JAN	Japanese Accepted Name
JMAF	Japanese Ministry for Agriculture, Forestry and Fisheries
K	temperature (Kelvin)
L	generalised ligand
LC	lethal concentration
LD	Lethal dose; LD_{50}: a dose which is lethal to 50% of the animals tested
M	relative molecular mass (formula weight)
M	metal
m	medium
mcd	magnetic circular dichroism
Me	methyl
MEL	maximum exposure limit
MeOH	methanol
mer	meridional
mes	mesityl (1,3,5-trimethylphenyl)
Me_2CO	acetone
misc.	miscible
misc.	miscellaneous
mixt.	mixture
mky	monkey
MOCVD	metal-organic chemical vapour deposition
mod.	moderately
Mp	melting point
mus	mouse
n	index of refraction eg. (n_D^{20} for 20° and sodium light).
nbd	norbornadiene
nqr	nuclear quadrupole resonance spectrum

Abbreviation	*Meaning*
obt.	obtained
oc	open cup
oep	octaethylporphyrinato
OES	occupational exposure standard
O_h	octahedral
op	optical purity
org.	organic
orl	oral
ox	oxalato
Ph	phenyl (C_6H_5)
pH	Measure of soln. acidity where pH = $\log_{10}(1/[H^+])$ where $[H^+]$ is the hydrogen ion
phen	1,10-phenanthroline
phys.	physical
p*K*	Measure of dissoc. const. (*K*) where p*K* = $\mathrm{Log}_{10}(1/K)$
pm	picometres (10^{-12} m)
PMDET	pentamethyldiethylenetriamine
polarog.	polarography
polym.	polymerised, polymerisation
ppm	parts per million
Pr	propyl (Pr^i for isopropyl)
prob.	probably
purifn.	purification
Py	pyridine
pz	pyrazolato
R	generalised alkyl group
rbt	rabbit
ref.	reference
rel.	relative(ly)
r.t.	room temperature
s	strong
S^0	standard entropy
scu	subcutaneous
skn	skin
sl.	slightly
sol.	soluble
soln(s)	solution(s)
solv(s)	solvent(s)
soly.	solubility
sp.	species (singular)
spar.	sparingly
spp.	species (plural)
ssp.	subspecies
subl.	sublimation, sublimes
tbp	triagonal bipyramidal
T_d	tetrahedral
Tf	triflate
THF	tetrahydrofuran
tht	tetrahydrothiophene
TLV	Threshold Limit Value
TMED	tetramethylethylenediamine
tpp	tetraphenylporphyrinato
triphos	triphosphine (generalised ligand)
Ts	tosyl

Abbreviation	*Meaning*
μ_{eff}	effective magnetic moment (in Bohr magnetons μ_B)
unsatd.	unsaturated
USAN	United States Adopted Name
uv	ultraviolet spectrum
v.	very
var.	variety
vis.	visible
vol.	volume
w	weak
WSSA	Weed Science Society of America
X	generalised anion, usually halide

Table 2. Reference tags

The following is a selection of the most common reference tags that are used

Abbreviation	*Meaning*
abs config	absolute configuration
anal	analysis
bibl	bibliography
biodistribn	biodistribution
biosynth	biosynthesis
cd	circular dichroism
chromatog	chromatography
cmr	^{13}C nuclear magnetic resonance spectrum
config	configuration
conformn	conformation
cryst struct	X-ray crystal structure determination
deriv(s)	derivative(s)
detn	determination, detection
dsc	differential scanning calorimetry
dta	differential thermal analysis
ed	electron diffraction
electrochem	electrochemistry, cyclic voltammetry
em	electron microscopy
epr	electron paramagnetic (spin) resonance spectrum
esca	electron spectroscopy for chemical analysis
exafs	extended X-ray diffraction fine structure
fab-ms	fast atom bombardment mass spectroscopy
glc	gas-liquid chromatography
haz	hazard
hplc	high performance liquid chromatogrpahy
ir	infrared spectrum
isol	isolation
isom	isomerism
manuf	manufacture
metab	metabolism
mineral	mineralogy
ms	mass spectrum
nmr	nuclear magnetic resonance spectrum
occur	occurrence

Abbreviation	*Meaning*
ord	optical rotatory dispersion
pe	photoelectron spectroscopy
pharmacol	pharmacology
photol	photoysis
pmr	proton (^{1}H) nuclear magnetic resonance spectrum
polarog	polarography
powder struct	X-ray powder structure determination
props	properties (chemical or physical)
Raman	Raman spectrum
resoln	resolution
rev	review
sepn	separation
soly	solubility
spectra	
struct	structure
synonyms	
synth	synthesis
tautom	tautomerism
tga	thermogravimetric analysis
theory	MO calculations etc.
tlc	thin layer chromatography
tox	toxicity
trans	transition(s)
use(s)	
uv	ultraviolet spectrum
uv-vis	ultraviolet visible spectrum